



Optimum and Equilibrium in Assignment Problems With Congestion: Mobile Terminals Association to Base Stations

Alonso Silva, Hamidou Tembine, Eitan Altman, Merouane Debbah

► To cite this version:

Alonso Silva, Hamidou Tembine, Eitan Altman, Merouane Debbah. Optimum and Equilibrium in Assignment Problems With Congestion: Mobile Terminals Association to Base Stations. IEEE Transactions on Automatic Control, 2013, 58 (8), pp.2018-2031. 10.1109/TAC.2013.2250072 . hal-00913217

HAL Id: hal-00913217

<https://inria.hal.science/hal-00913217>

Submitted on 10 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimum and equilibrium in assignment problems with congestion: mobile terminals association to base stations

Alonso Silva, Hamidou Tembine,
Eitan Altman, Mérouane Debbah

Abstract

The classic optimal transportation problem consists in finding the most cost-effective way of moving masses from one set of locations to another, minimizing its transportation cost. The formulation of this problem and its solution have been useful to understand various mathematical, economical, and control theory phenomena, such as, e.g., Witsenhausen's counterexample in stochastic control theory, the principal-agent problem in microeconomic theory, location and planning problems, etc. In this work, we incorporate the effect of network congestion to the optimal transportation problem and we are able to find a closed form expression for its solution. As an application of our work, we focus on the mobile association problem in cellular networks (the determination of the cells corresponding to each base station). In the continuum setting, this problem corresponds to the determination of the locations at which mobile terminals prefer to connect (by also considering the congestion they create) to a given base station rather than to other base stations. Two types of problems have been addressed: a global optimization problem for minimizing the total power needed by the mobile terminals over the whole network (global optimum), and a user optimization problem, in which each mobile terminal chooses to which base station to connect in order to minimize its own cost (user equilibrium). This work combines optimal transportation with strategic decision making to characterize both solutions.

I. INTRODUCTION

Optimal transportation problems have their origins in planning problems (e.g., optimally transporting coal from mines to steel factories). In the classic optimal transportation problem (centralized scenario), a central planner has to find some transportation plan between two nonnegative probability measures so as to minimize the average transportation cost. Such problems have been very much studied in recent years for their various applications in, for example, partial differential equations, control theory, probability theory, microeconomic theory, etc. Typical applications of these problems are optimal allocation of resources or assignment problems that are frequently met in economics as well as in engineering.

The most common cost functions considered in optimal transportation depend basically on some distance between two probability measures (Monge-Kantorovich distance, Lévy-Prokhorov metric, Wasserstein distance, total variation distance, etc.). However, in many situations the decision of which assignment to choose not only depends on the distance but it heavily depends on the congestion that the assignment generates. For example, one of many cases where this situation arises is in the mobile association problem in cellular networks. This problem is basically an assignment problem and it consists on the determination of cells corresponding to each base station of the network. In the continuum setting, this problem corresponds to the determination of the locations at which mobile terminals prefer to connect (by also considering the congestion they create) to a given base station rather than to other base stations. We consider the case where mobile terminals within a cell share the same spectrum, and consequently, mobile terminals decisions to which base station to connect affect the decision making process of other mobile terminals in the network due to network congestion. From these interactions, mobile terminals learn their optimal access point until they reach an equilibrium, where the user optimization (equilibrium) depends upon the context (type of service, bit rate, delay, etc.). Within this context, two types of problems

Part of this work has been presented at the 49th IEEE Conference on Decision and Control.

Alonso Silva is with UC Berkeley. Email: asilva@eecs.berkeley.edu

Eitan Altman is with INRIA. Email: eitan.altman@inria.fr

Hamidou Tembine and Mérouane Debbah are with Supélec. Email: Name.Surname@supelec.fr

have been considered: a global optimization problem for minimizing the total power needed to satisfy a certain throughput for the mobile terminals of the network (global optimum), and a user optimization (equilibrium), in which each mobile terminal chooses to which base station to connect in order to minimize its own expected cost.

From a game theory perspective, in the decentralized scenario, the assignment problem consists on each mobile terminal (player) choosing the base station that best serve its needs. This game corresponds to an anonymous game since each player's utility depends on how many of his peers choose which strategy (in this particular case, the strategy of each player corresponds to which base station to connect). Roughly speaking, this is translated into the fact that the utility function is invariant by permutation of the index of the players. This property is sometimes referred as exchangeability or indistinguishability property.

Starting from the seminal paper of Hotelling [1] a large area of research on location games has been developed. In [1], the author introduced the notion of spatial competition in a duopoly situation. Plastria [2] presented an overview of the research on locating one or more new facilities in an environment where competing facilities already exist. Gabszewicz and Thisse [3] provided another general survey on location games. Altman et al. [4] studied the duopoly situation in the uplink scenario of a cellular network where the users are placed on a line segment. The authors realized that, considering the particular cost structure that arises in the cellular context, complex cell shapes are obtained at the equilibrium. Other works incorporating congestion in optimal transportation are [5] and [6]. In [5], the authors incorporate the Wardrop equilibrium in a congested network by assigning probabilities to the used routes. The work of [6] is more related to our work where the authors consider an additive congestion cost on the objective function. Our work focuses on the additive and multiplicative congestion cost on the objective function, and the Wardrop equilibrium in the decentralized scenario. As an application of our work, we consider the mobile association problem in the downlink scenario (transmission from base stations to mobile terminals) and in a more general situation in a one-dimensional and two-dimensional scenarios without making any assumption on the symmetry of the users location. In order to do that, we propose a framework for the mobile association problem using optimal transportation (see the recent books of Villani [7], [8] and references therein). This theory was pioneered by Monge [9] and Kantorovich [10] and it has been proven to be useful in many mathematical, economical, and control theory contexts (see, e.g., [11], [12], [13]). There is a number of works on "optimal transportation" (see [14], and references therein) however the authors in [14] consider only an optimal selection of routes but do not use the rich theory of optimal transportation. The works on stochastic geometry are similar to our analysis of wireless networks (see, e.g., [15] and references therein) but in our case we do not consider any particular deployment distribution function. Note that the optimization aspect is not especially exploited in [15]. Fluid models allow one to have this general deployment distribution function.

Optimal transportation is a general framework that allows one to better understand the deployment of large-scale congested networks in a more efficient way. This framework also explains the cell-formation and user distribution over the wireless network.

In summary, in the current work, we study the centralized and decentralized optimal transportation problem by incorporating the congestion that the assignment generates over the network. As an application of our work, we focus on the minimization of the total power of the network in both the centralized and decentralized mobile association problems, while maintaining a certain level of throughput for each user connected to the network (this focus can be considered as an energy-efficient objective). We propose the rich theory of optimal transportation as the main tool of modelization of these mobile association problems. Thanks to this approach, we are able to characterize these mobile associations under different policies and give illustrative examples of this technique.

The remaining of this paper is organized as follows. In Section II we give some basic concepts and results in optimal transportation. Then we incorporate the congestion in this setting and give closed form expressions for its solution. Section III outlines the problem formulation for minimizing the total power of a wireless network under quality of service constraints (to maintain a certain level of throughput for each user connected to the network). We address the problem for the downlink case. Two different policies

are considered: round robin scheduling policy (also known as time fair allocation policy) and rate fair allocation policy. Round robin scheduling policy and rate fair allocation policy are studied in detail in Section IV and Section V respectively. In Section VI we study the price of anarchy or performance gap between the centralized and decentralized scenarios. In Section VII we give numerical examples for one-dimensional and two-dimensional mobile terminals deployment distribution functions. Section VIII concludes this work.

II. BASICS IN OPTIMAL TRANSPORTATION

Optimal transportation, also called theory of mass transportation, goes back to the original works by Monge in 1781 [9], and later in 1942 by Kantorovich [10]. The work of Brenier in 1987 [16] has renewed the interest for the subject and since then many different works have been written in this topic (see e.g. the recent books of Villani [7], [8] and references therein).

The original problem of Monge was:

“What is the optimal way to move piles of sand
to fill up given holes of the same total volume?”

New approaches to formulate and to tackle this problem have lead to applications in many different fields such as partial differential equations, microeconomics, control theory, etc. Through this work, we will see how to incorporate the network congestion in this problem and how its solution can be applied to cellular networks.

In uplink transmissions in cellular networks, we have mobile terminals sending information to base stations and we are interested to know the optimal way to “move” information from mobile terminals to base stations. As we will see, it turns out that even if both questions are quite different, they also share some similarities. Through this section we will refer to this example to fix some ideas and it will also help us to the development of the following sections.

The mathematical framework for this problem is the following: we consider a domain (connected open subset of a finite-dimensional vector space), denoted by \mathcal{D} , two probability measures μ and ν on that domain, defined over a probability space $(\mathcal{D}, \mathcal{F}, \mathbb{P})$.

For example, assume that the probability measure μ has a density function f , and that the probability measure ν has a density function g ¹. The density function f could represent the density function of the sandpiles in the one-dimensional space. The density function g could represent the density function of the holes in the one-dimensional space.

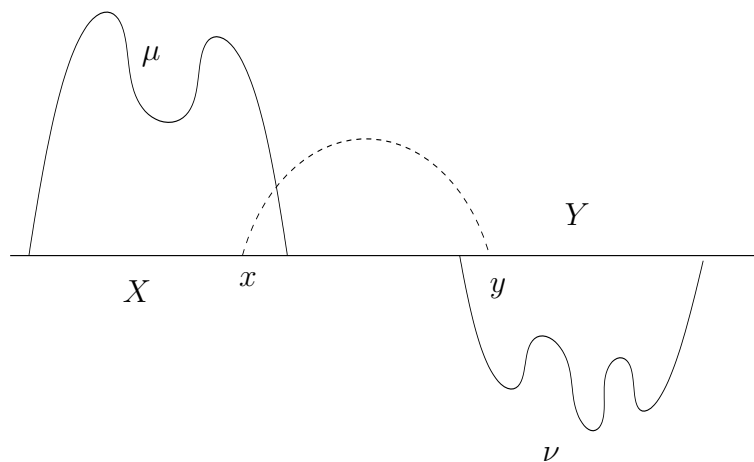


Fig. 1: Monge's problem.

¹Not every probability measure has a density function. The condition to have a probability density function is that the probability measure to be absolutely continuous with respect to Lebesgue measure.

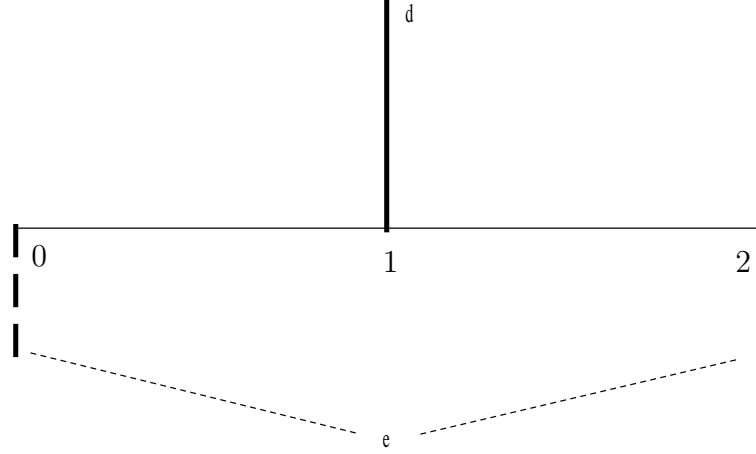


Fig. 2: Monge's problem can not model a simple scenario of two mobile terminals and one base station. Kantorovich's problem however can model very general scenarios.

A function T which relates both probability density functions is called a transport map if the following condition is satisfied:

$$\int_A F(y)g(y) dy = \int_{\{x \in X : T(x) \in A\}} F(T(x))f(x) dx, \quad (1)$$

for all continuous function F , where X is the support² of the probability density function f and we denote this condition (following the optimal transportation notation) as

$$T\#\mu = \nu. \quad (2)$$

In the previous example, the transport map would be the transfer of information from location s to location t . It basically associates mobile terminals to base stations and transports information between them. The condition (1) is an equation of conservation of the information.

Notice that, in communication systems there exists packet loss so in general this condition may never be satisfied, but considering an estimation of the packet loss (for example, by sending standard packets test), this condition can be satisfied by modifying the reception measure ν . If we can not obtain a good estimation of this reception measure, we can consider it in its current form as a conservative policy.

In the original problem, Monge considered that the cost of moving a commodity from position x to a position y depends on the distance $|x - y|$, i.e., the cost function $c \equiv c(|x - y|)$. The cost of moving a commodity from position x through T to its new position $T(x)$ will then be given by $c(|x - T(x)|)$. For the global optimization problem, we need to consider the additive total cost over the network multiplied by the quantity of information that it is being transferred, which in the continuum setting will be:

$$\text{Min} \int_{\mathcal{D}} c(|x - T(x)|) f(x) dx \quad \text{such that} \quad T\#\mu = \nu, \quad (3)$$

where μ and ν are probability measures and $T : \mathcal{D} \rightarrow \mathcal{D}$ is an integrable function. This problem is known as Monge's problem in optimal transportation.

Not only Monge's problem is a difficult problem because of the highly non-linear structure of its objective function, but it also presents several limitations in its formulation. Examples of the limitations in Monge's formulation can be found in [7]. As a simple example, consider the interval $\mathcal{D} = [0, 2]$, the probability function $\mu = \delta_1$ (the Dirac delta function concentrated at 1)³ and the probability function $\nu =$

²The support of a function f is the closure of the set of points where the function is not zero, i.e., $\text{support}(f) = \overline{\{t : f(t) \neq 0\}}$

³The Dirac delta function concentrated at point a can be thought of as a function such that

$$\delta_a(x) = \begin{cases} +\infty, & x = a \\ 0, & x \neq a \end{cases} \quad (4)$$

$\frac{1}{2}(\delta_0 + \delta_2)$ (the semisum of two Dirac delta functions one concentrated at 0 and the other one concentrated at 2). Since in Monge's formulation, there is no splitting of mass, i.e., everything that is transmitted from one location x has to go to another location $T(x)$, this simple problem doesn't have a transport map (see Fig. 3). This limitation is due in part to the original considered problem, but as we will see this limitation is overcome in Kantorovich's formulation. Kantorovich considered another formulation of this problem in [10].

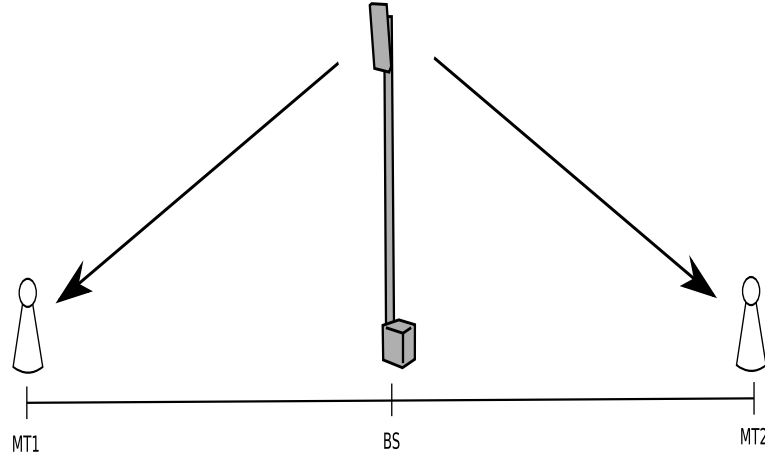


Fig. 3: Monge's problem can not model a simple scenario of two mobile terminals and one base station. Kantorovich's problem however can model very general scenarios.

Kantorovich noticed that the problem of transportation from one location to another can be seen as "graphs of functions" (called transport plans) in the product space (See Fig. 4).

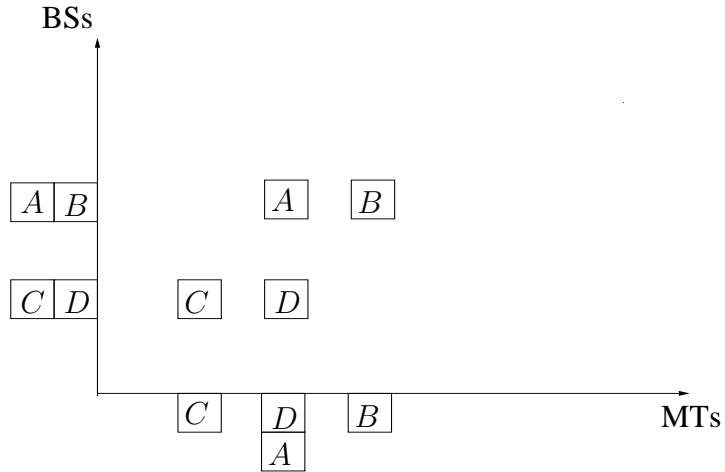


Fig. 4: Kantorovich considered "graphs" where the projection in the first axis coincide with the mobile terminal position ($MT_1 = 3.5$, $MT_2 = 5$ and $MT_3 = 6.5$) and the second axis coincides with the base station position ($BS_1 = 4$ and $BS_2 = 6$).

The idea is to minimize the objective function over the space of graphs $g = (\text{Id} \times T)$ in the product space. Then with the condition that each mobile terminal satisfies its uplink demand and that the information is

constrained to satisfy the identity

$$\int_{-\infty}^{+\infty} \delta_a(x) dx = 1. \quad (5)$$

This is a heuristic definition. However the Dirac delta function can be rigorously defined either as a distribution or as a measure. For a rigorous definition, see [17].

received at the base stations, Kantorovich's problem reads

$$\text{Min}_{g \in \Pi(\mu, \nu)} \iint_{\mathcal{D} \times \mathcal{D}} c(x, y) dg(x, y), \quad (6)$$

where

$$\Pi(\mu, \nu) = \{g : \pi_1 \# g = \mu \text{ and } \pi_2 \# g = \nu\}, \quad (7)$$

is denoted the ensemble of transport plans g , $\pi_1(x, y)$ stands for the projection on the first axis x , and $\pi_2(x, y)$ stands for the projection on the second axis y .

The relationship between Monge and Kantorovich problems is that every transport map T of Monge's problem determines a transport plan $g_T = (\text{Id} \times T) \# \mu$ in Kantorovich's problem with the same cost (where Id denotes the identity). However, Kantorovich's problem considers more functions than the ones coming from Monge's problem (which can always be viewed as the product of the identity and the map T), so we can choose from a bigger set $\Pi(\mu, \nu)$.

Then, every solution of Kantorovich's problem is a lower bound to Monge's problem, *i.e.*,

$$\text{Min}_{g \in \Pi(\mu, \nu)} \iint_{\mathcal{D} \times \mathcal{D}} c(x, y) dg(x, y) \leq \text{Min}_{T \# \mu = \nu} \int_{\mathcal{D}} c(|x - T(x)|) f(x) dx. \quad (8)$$

The Theorem below provides an equivalence result between the two problems for the cost function $|x - y|^p$, $p > 1$ and a continuous measure.

Theorem 2.1: Consider the cost function $c(|x - y|) = |x - y|^p$. Let μ and ν be probability measures in \mathcal{D} and fix $p \geq 1$. We assume that μ can be written⁴ as $d\mu = f(x) dx$. Then the optimal value of Monge's problem coincides with the optimal value of Kantorovich's problem, *i.e.*, $M_p(\mu, \nu) = W_p(\mu, \nu)$ and there exists an optimal transport map from μ to ν , which is also unique almost everywhere if $p > 1$.

This result is very difficult to obtain and it has been proved only recently (see [16] for the case $p = 2$, and the references in [7] for the other cases).

The case that we are interested in can be characterized because the image of the transport plan is a discrete finite set.

Thanks to optimal transport theory we are able to characterize the partitions considering general settings. To this purpose, consider locations $(x_1, y_1) \dots, (x_K, y_K)$, the Euclidean distance $d_i(x, y) = \sqrt{(x - x_i)^2 + (y - y_i)^2}$, and F a continuous function.

Theorem 2.2: Consider the problem

$$(P1) \quad \text{Min}_{C_i} \sum_{i=1}^K \iint_{C_i} \left[F(d_i(x, y)) + s_i \left(\iint_{C_i} f(\omega, z) d\omega dz \right) \right] f(x, y) dx dy, \quad (9)$$

where C_i is the cell partition of \mathcal{D} . Suppose that s_i are continuously differentiable, non-decreasing, and convex functions. The problem (P1) admits a solution that verifies

$$(S1) \quad \begin{cases} C_i = \{(x, y) : F(d_i(x, y)) + s_i(N_i) + N_i \cdot s'_i(N_i) \\ \leq F(d_j(x, y)) + s_j(N_j) + N_j \cdot s'_j(N_j)\} \\ N_i = \iint_{C_i} f(\omega, z) d\omega dz. \end{cases} \quad (10)$$

Proof.- See Appendix A

Note that the system (10) provides a fixed-point system between C_i and N_i , and can be interpreted as follows: the distribution of users in C_i , generates the number N_i and which affects the congestion term via the function $s_i(\cdot)$, which in consequence influences the set C_i .

⁴The exact condition is that μ is absolutely continuous with respect to the Lebesgue measure. A probability measure μ is absolutely continuous with respect to the Lebesgue measure if the function $F(x) = \mu((-\infty, x])$ is locally an absolutely continuous real function. A function f is an absolutely continuous real function if there exists an integrable function g such that $f(x) = f(a) + \int_a^x g(t) dt$

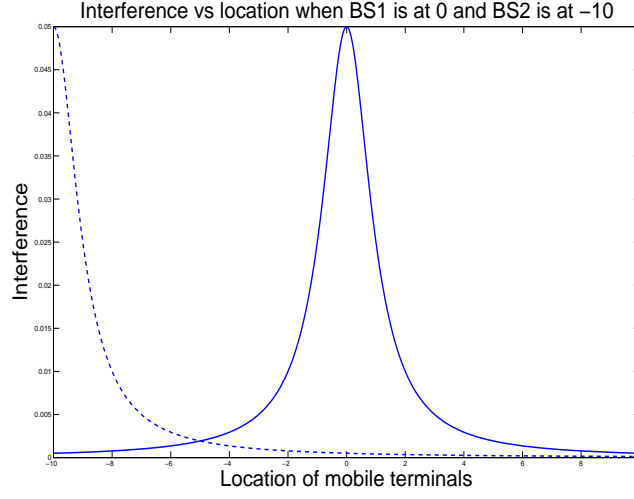


Fig. 5: Interference as a function of location of mobile terminals when BS₁ is at position 0 (solid line) and BS₂ at -10 (dotted line).

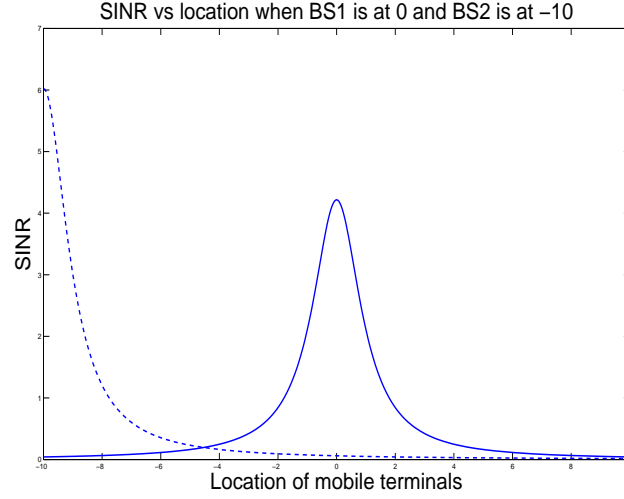


Fig. 6: SINR as a function of location of mobile terminals when BS₁ is at position 0 (solid line) and BS₂ at -10 (dotted line).

Theorem 2.3: Consider the problem

$$(P2) \quad \text{Min}_{C_i} \sum_{i=1}^K \iint_{C_i} \left[F(d_i(x, y)) \cdot m_i \left(\iint_{C_i} f(\omega, z) d\omega dz \right) \right] f(x, y) dx dy, \quad (11)$$

where C_i is the cell partition of \mathcal{D} . Suppose that m_i are derivable. The problem (P2) admits a solution that verifies

$$(S2) \quad \begin{cases} C_i = \{(x, y) : m_i(N_i)F(d_i(x, y))f(x, y) + U_i(x, y) \\ \leq m_j(N_j)F(d_j(x, y))f(x, y) + U_j(x, y)\} \\ U_i = m'_i(N_i) \iint_{C_i} F(d_i(x, y))f(x, y) dx dy \\ N_i = \iint_{C_i} f(\omega, z) d\omega dz. \end{cases} \quad (12)$$

Proof.- See Appendix B

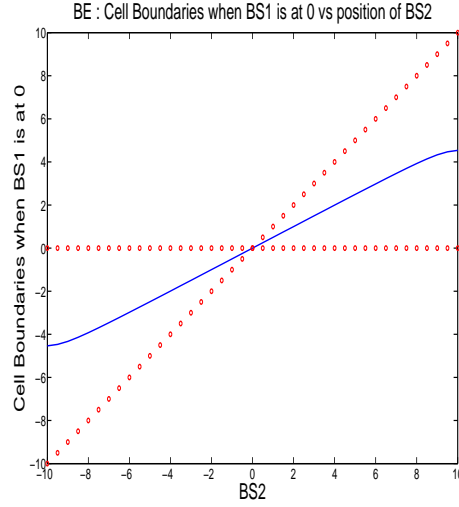


Fig. 7: Wardrop Equilibrium: Thresholds determining the cell boundaries as a function of the location of the base stations. The network is deployed over the interval $[-10, 10]$ (one-dimensional case) presented here vertically. We consider a uniform distribution of mobile terminals and we find the threshold (solid line) determining the cell boundaries as a function of the base stations positions (dotted lines) by changing the position of one of them. BS_1 is fixed at position 0 and we change the position of BS_2 from -10 to $+10$.

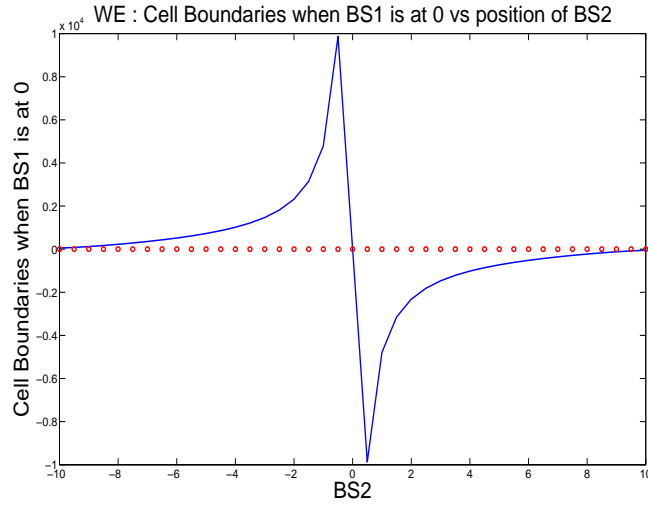


Fig. 8: Worst Equilibrium: Thresholds determining the cell boundaries (vertical axis) that give the worst equilibrium in terms of the SINR as a function of the location of BS_2 for BS_1 at position 0.

Notice that in problem (P1) if the functions $s_i \equiv 0$ the solution of the system (S1) becomes the well known Voronoi cells. In problem (P2) if we have that the functions $h_i \equiv 1$ we find again the Voronoi cells. In general however, the Voronoi configuration is far from being optimal.

III. THE SYSTEM MODEL AND PROBLEM FORMULATION

A summary of the notation used on this work can be found in Table I. We consider a network deployed on a region, denoted by \mathcal{D}^5 , over the two-dimensional plane. The mobile terminals (MTs) are distributed

⁵The domain \mathcal{D} is considered bounded.

TABLE I: Notation

N	Total number of mobile terminals in the network
K	Total number of BSs
f	Deployment distribution of mobile terminals
(x_i, y_i)	Position of the i -th BS
C_i	Cell determined by the i -th BS
N_i	Number of mobile terminals associated to the i -th BS
MAX_i	Number of carriers offered by the i -th BS
κ_i	Penalization function of non-service
h_i	Channel gain function in the i -th cell
ξ_i	Path loss exponent in the i -th cell

according to a given probability density function $f(x, y)$. The proportion of mobile terminals in the sub-region $\mathcal{A} \subseteq \mathcal{D}$ is given by

$$\iint_{\mathcal{A}} f(x, y) dx dy. \quad (13)$$

The number of mobile terminals in a sub-region of the network $A \subseteq \mathcal{D}$, denoted by $N(A)$, can be approximated by

$$N(A) = N \left(\iint_A f(x, y) dx dy \right), \quad (14)$$

where N is the total number of mobile terminals in the network. The integral on the right hand side between brackets takes into account the proportion of mobile terminals distributed over the network area A . The equation 14 does not give necessarily an integer. We will take integer part (floor) if needed. Examples of mobile terminals distributions $f(x, y)$:

- 1) If the mobile terminals are distributed uniformly over the network, then the probability density function of the mobile terminals is given by

$$f(x, y) = \frac{1}{|\mathcal{D}|}, \quad (15)$$

where $|\mathcal{D}|$ is the total area of the network.

- 2) If the mobile terminals are distributed according to different levels of population density, then the probability density function of the mobile terminals is given by

$$f(x, y) = \begin{cases} f_{\text{HD}} & \text{if } (x, y) \text{ is at a High Density region,} \\ f_{\text{ND}} & \text{if } (x, y) \text{ is at a Normal Density region,} \\ f_{\text{LD}} & \text{if } (x, y) \text{ is at a Low Density region,} \end{cases} \quad (16)$$

where $f_{\text{HD}} > f_{\text{ND}} > f_{\text{LD}}$ are defined similarly to equation (15) with constants of normalization K_{HD} , K_{ND} , K_{LD} , such that $K_{\text{HD}} > K_{\text{ND}} > K_{\text{LD}}$.

- 3) If the mobile terminals are distributed in a circle with more mobile terminals in the center of the network area and less mobile terminals towards the boundary, then we can consider that the probability density function of the mobile terminals is given by

$$f(x, y) = \frac{R_D^2 - (x^2 + y^2)}{K_D}, \quad (17)$$

where R_D is the radius of the network and K_D is a coefficient of normalization.

In the network, we consider K base stations (BSs), denoted by $BS_i, i \in \{1, \dots, K\}$, located at fixed positions $(x_i, y_i), i \in \{1, \dots, K\}$. We assume that the neighbouring BSs transmit their signals in orthogonal frequency bands. Furthermore, we assume that interference between BSs that are far from each other is negligible. Consequently, instead of considering the SINR (Signal to Interference plus Noise Ratio), we consider as performance measure the SNR (Signal to Noise Ratio). We consider the downlink case (transmission from base stations to mobile terminals) and assume that each BS is going to transmit only to mobile terminals associated to it. We denote by C_i the set of mobiles associated to the i -th BS, and by N_i the number of mobiles within that cell, i.e., the cardinal of the set C_i , to be determined.

Since the probability density function of the mobile terminals $f(x, y)$ considered in our work is general, instead of considering a particular probability density function, e.g., $f(x, y)$, and an average throughput, e.g., $\tilde{\theta}(x, y)$, in each location (x, y) , we can consider a constant throughput $\theta > 0$ and redefine the distribution of mobile terminals $f(x, y)$ as

$$f(x, y) := \frac{\tilde{f}(x, y)\tilde{\theta}(x, y)}{\theta}. \quad (18)$$

In order to do that, we notice that since $f(x, y)$ must be a probability density function, we should impose

$$\iint f(x, y) dx dy = 1. \quad (19)$$

This equation is equivalent to

$$\iint \frac{\tilde{f}(x, y)\tilde{\theta}(x, y)}{\theta} dx dy = 1. \quad (20)$$

Since we are considering a constant throughput $\theta > 0$, this equation is equivalent to

$$\frac{1}{\theta} \iint \tilde{f}(x, y)\tilde{\theta}(x, y) dx dy = 1. \quad (21)$$

For this equation to hold, we consider

$$\theta := \iint \tilde{f}(x, y)\tilde{\theta}(x, y) dx dy. \quad (22)$$

In consequence, the following equation holds:

$$f(x, y)\theta = \tilde{f}(x, y)\tilde{\theta}(x, y). \quad (23)$$

Last equation simply translates in the fact that, e.g., mobile terminals with double demand than others would be considered as two different mobile terminals with the same demand than other mobile terminals at the same location.

If the number of mobile terminals is greater than the maximum number of carriers available in the i -th cell, denoted by MAX_i , we consider a penalization cost function given by

$$\kappa_i(N_i) = \begin{cases} 0 & \text{if } N_i \leq MAX_i, \\ \bar{\kappa}_i(N_i - MAX_i) & \text{if } N_i > MAX_i. \end{cases} \quad (24)$$

We assume that $\bar{\kappa}_i$ can be either a constant function or at least a non-decreasing function in the number of used carriers above the limit MAX_i ⁶. We first study the case $N_i \leq MAX_i$ and we study the general case in Section V-A.

The power transmitted P^T from a base station BS_i to a mobile terminal located at position (x, y) is denoted by

$$P_i^T(x, y) = P_i(x, y). \quad (25)$$

⁶For example, the maximum number of possible carriers in WiMAX is around 2048, so by using this technology we have $MAX_i = 2048$.

The received power P^R at a mobile terminal located at position (x, y) served by a base station BS_i is given by

$$P_i^R(x, y) = P_i(x, y)h_i(x, y), \quad (26)$$

where $h_i(x, y)$ is the channel gain between base station BS_i and a mobile terminal located at position (x, y) . We shall further assume that the channel gain corresponds to the path loss given by

$$h_i(x, y) = (\sqrt{R^2 + d_i^2(x, y)})^{-\xi}, \quad (27)$$

where ξ is the path loss exponent, R is the height of the base station, and $d_i(x, y)$ is the distance between a mobile terminal at position (x, y) and BS_i located at (x_i, y_i) , *i.e.*,

$$d_i(x, y) = \sqrt{(x_i - x)^2 + (y_i - y)^2}. \quad (28)$$

The SNR received at mobile terminals at position (x, y) in cell C_i is given by

$$\text{SNR}_i(x, y) = \frac{P_i(x, y)h_i(x, y)}{\sigma^2}, \quad (29)$$

where σ^2 is the noise power. We assume that the instantaneous mobile throughput is given by the following expression, which is based on Shannon's capacity theorem:

$$\theta_i(x, y) = \log(1 + \text{SNR}_i(x, y)). \quad (30)$$

We want to satisfy an average throughput for mobile terminals located at position (x, y) given by $\tilde{\theta}(x, y) > 0$. We shall consider for this objective two policies defined in [18]:

- (A) *Round robin scheduling policy*: where each base station devotes an equal fraction of time for transmission to each mobile terminal associated to it, and
- (B) *Rate fair allocation policy*: where each base station maintains a constant power sent to mobile terminals within its cell but modifies the fraction of time allowed to mobile terminals with different channel gains, so that the average transmission rate demand is satisfied.

For more information about this type of policies in the one dimensional case, see [18].

A. Round robin scheduling policy: Global Optimization

Following this policy, a base station devotes an equal fraction of time for transmission to mobile terminals located within its cell. The set of mobile terminals associated to the i -th base station BS_i , denoted by C_i , and the number of mobile terminals within that cell N_i (which is equal to the cardinal of the set C_i) are to be determined. Since a base station BS_i divides its time of service proportional to the number of mobile terminals within its cell, then the throughput of a mobile terminal located at position (x, y) under the round robin scheduling policy is given by

$$\theta_i^{\text{RR}}(x, y) = \frac{1}{N_i}\theta_i(x, y), \quad (31)$$

where N_i is the number of mobile terminals within the i -th cell and $\theta_i(x, y)$ is the instantaneous mobile throughput at mobile terminals located at positions (x, y) from base station BS_i . In order to satisfy an instantaneous throughput $\tilde{\theta}(x, y)$, we need to have

$$\theta^{\text{RR}}(x, y) = \frac{1}{N_i}\theta_i(x, y) \geq \tilde{\theta}(x, y), \quad (32)$$

or equivalently, by replacing in the previous equation $\theta_i(x, y)$ from equations (29) and (30),

$$\frac{1}{N_i} \log \left(1 + \frac{P_i(x, y)h_i(x, y)}{\sigma^2} \right) \geq \tilde{\theta}(x, y). \quad (33)$$

Rewriting last equation, we obtain

$$P_i(x, y) \geq \frac{\sigma^2}{h_i(x, y)} (2^{N_i \bar{\theta}(x, y)} - 1). \quad (34)$$

where $h_i(x, y) > 0$. Since our objective function is to minimize the total power of the network, this constraint will be reached, and from equation (27) we obtain

$$P_i(x, y) = \sigma^2 (2^{N_i \bar{\theta}(x, y)} - 1) ([R^2 + d_i^2(x, y)]^{1/2})^{+\xi}. \quad (35)$$

From equation (35), we remark that:

- (a) If the number of mobile terminals increases within a cell, i.e., the base station has more mobile terminals assigned to it, the base station will need to transmit more power to each of the mobile terminals within its cell, to satisfy the throughput constraint in equation (32).
- (b) The function $([R^2 + d_i^2(x, y)]^{1/2})^{+\xi}$ on the right hand side give us the dependence of the power needed to satisfy the throughput constraint in equation (32) with respect to the distance between the base station and the mobile terminal located at position (x, y) .

Our objective is to find the optimal mobile association in order to minimize the total power of the network while maintaining the throughput constraint given by equation (32), or equivalently by equation (35). The total power of the network is equal to the sum of the power used in each cell in the network, i.e.,

$$P_{\text{total}} = \sum_{i=1}^K P_i^{\text{intra}}, \quad (36)$$

$$\text{where } P_i^{\text{intra}} = \iint_{C_i} P_i(x, y) f(x, y) dx dy, \quad (37)$$

P_i^{intra} is the intracell power consumption in cell C_i .

The global optimization for the mobile association problem is to determine the cells C_i , $i \in \{1, \dots, K\}$, to minimize the total power of the network:

$$(\text{RR}) \quad \min \sum_{i=1}^K \iint_{C_i} P_i(x, y) f(x, y) dx dy, \quad (38)$$

subject to (35), where $f(x, y)$ is the probability density function of the mobile terminals. We solve this problem in Section III-A.

B. Rate fair allocation policy: User Optimization

In the rate fair allocation policy, each base station will maintain a constant power sent to mobile terminals within its cell, i.e.,

$$P_i(x, y) = P_i \text{ for each mobile terminal at location } (x, y) \text{ inside cell } C_i. \quad (39)$$

However, the base station modifies the fraction of time allotted to mobile terminals, set in such a way that the average transmission rate to each mobile terminal with different channel gain is the same, denoted by $\tilde{\theta}(x, y)$, for each mobile terminal located at position (x, y) .

Let r_i be the fixed rate of mobile terminals located inside the cell C_i . The fraction of time that a mobile terminal located at position $(x, y) \in C_i$ receives positive throughput is

$$\frac{r_i}{\log(1 + \text{SNR}_i(x, y))}. \quad (40)$$

Then, the fixed rate r_i is the solution to equation

$$\iint_{C_i} \frac{r_i}{\log(1 + \text{SNR}_i(x, y))} f(x, y) dx dy = \beta, \quad (41)$$

where β is the fraction of time guaranteed from the regulator to have no interference from the other BSs. Then, the rate

$$r_i = \left(\iint_{C_i} \frac{1}{\log(1 + \text{SNR}_i(x, y))} f(x, y) dx dy \right)^{-1} \beta. \quad (42)$$

We study the equilibrium states where each mobile terminal chooses the base station which will serve it, i.e., the situation where given the interactions with other mobile terminals, each mobile terminal doesn't have any incentive to unilaterally change its strategy. A similar notion of equilibrium has been studied in the context of large number of small players in road-traffic theory by Wardrop [19]. A Wardrop equilibrium is the analog of a Nash equilibrium for the case of a large number of small players. In the following, we give an equilibrium notion that we also denote Wardrop equilibrium by analogy to the notion given in [19].

Definition.- The Wardrop equilibrium is given in the context of cellular systems by:

$$\text{If } \iint_{C_i} f(x, y) dx dy > 0, \text{ then } \theta_i = \max_{1 \leq j \leq K} \theta_j(C_j), \quad (43a)$$

$$\text{else if } \iint_{C_i} f(x, y) dx dy = 0, \text{ then } \theta_i \leq \max_{1 \leq j \leq K} \theta_j(C_j). \quad (43b)$$

As we previously mentioned, a Wardrop equilibrium is the analog of a Nash equilibrium in the case of a large number of small players, where, in our case, we consider the mobile terminals as the small players. In this setting, the Wardrop equilibrium indicates that if there is a strictly positive proportion of mobile terminals associated to the i -th base station (the left-hand side condition in equation (43a)), then the throughput that the mobile terminals associated to the i -th base station obtain is the maximum that they would obtain from any other base station (right-hand side consequence in equation (43a)). The second condition indicates that if there is one base station that doesn't have any mobile terminal associated to it (left-hand side condition in equation (43b)), it is because the mobile terminals can obtain a higher throughput by connecting to one of the other base stations (right-hand side consequence in equation (43b)).

We assume that each base station is serving at least one mobile terminal, (if that is not the case, we could remove the base stations which are not serving any mobile terminal and find an equivalent situation). Then, the equilibrium is given by the situation when

$$\theta_1 = \theta_2 = \dots = \theta_K. \quad (44)$$

To understand this equilibrium situation, consider as an example the simple case of two base stations: BS₁ and BS₂. Assume that the base station BS₁ offers more throughput than the base station BS₂. Then, the mobile terminals being served by the base station BS₂ will have an incentive to connect to the base station BS₁. From equation (35), we previously remarked that the transmitted throughput depends inversely on the number of mobile terminals connected to the base station. As more mobile terminals try to connect to base station BS₁ the throughput of each mobile terminal will diminish until arriving to the equilibrium where both base stations will offer the same throughput.

The condition given by equation (35) is equivalent in our setting to the condition

$$\text{SNR}_1 = \text{SNR}_2 = \dots = \text{SNR}_K. \quad (45)$$

Let us denote by β the SNR offered by the base stations at the equilibrium, i.e.,

$$\beta = \text{SNR}_1 = \text{SNR}_2 = \dots = \text{SNR}_K. \quad (46)$$

From equation (29) we know that

$$P_i(x, y) = \sigma^2 \frac{\text{SNR}_i(x, y)}{h_i(x, y)}. \quad (47)$$

with $h_i(x, y) > 0$.

At the equilibrium, from equations (46) and (29) we obtain

$$P_i(x, y) = \sigma^2 \frac{\beta}{h_i(x, y)}. \quad (48)$$

Replacing the term $P_i(x, y)$ in equation (37)

$$P_i^{\text{intra}} = \iint_{C_i} \sigma^2 \frac{\beta}{h_i(x, y)} f(x, y) dx dy, \quad (49)$$

Replacing the channel gain $h_i(x, y)$ from equation (27), we obtain

$$P_i^{\text{intra}} = \beta \iint_{C_i} \sigma^2 ([R^2 + d_i^2(x, y)]^{1/2})^{+\xi} f(x, y) dx dy. \quad (50)$$

We want to choose the optimal mobile assignment in order to minimize the total power of the network under the constraint that the mobile terminals have an average throughput of θ , *i.e.*,

$$\text{Min}_{C_i} P_{\text{total}} = \sum_{i=1}^K P_i^{\text{intra}}. \quad (51)$$

Then our problem reads

$$(\text{RF}) \quad \text{Min}_{C_i} \sum_{i=1}^K \iint_{C_i} \beta \sigma^2 (R^2 + d_i^2(x, y))^{\xi/2} f(x, y) dx dy. \quad (52)$$

We will solve this problem in Section V. Thanks to optimal transport theory we are able to characterize the partitions considering a general setting. In the following section, we will briefly describe optimal transport theory and motivate the solution of the previously considered mobile association problems.

IV. ROUND ROBIN SCHEDULING POLICY

We assume that a service provider wants to minimize the total power of the network while maintaining a certain average throughput of θ to each mobile terminal of the system using the round robin scheduling policy given by problem

$$(\text{RR}) \quad \text{Min}_{C_i} \sum_{i=1}^K \iint_{C_i} \sigma^2 (R^2 + d_i(x, y)^2)^{\xi/2} (2^{N_i\theta} - 1) f(x, y) dx dy. \quad (53)$$

We see that this problem is an optimal transportation problem like the one in (P1) with cost function given by

$$F(d_i(x, y)) = \sigma^2 (R^2 + d_i(x, y)^2)^{\xi/2} \quad \text{and} \quad (54)$$

$$m_i(x, y) = (2^{N_i\theta} - 1). \quad (55)$$

From the previous theorem, we can derive an explicit expression for this configuration.

Proposition.- There exist a unique optimum given by

$$C_i = \left\{ x_0 \in \mathcal{D} : d_i(x_0, y_0)^p + h_i(N_i) + N_i h'_i(N_i) \leq d_j(x_0, y_0)^p + k_j(N_j) + N_j k'_j(N_j), \forall j \neq i \right\} \quad (56)$$

$$N_i = \iint_{C_i} f(x_0, y_0) dx_0 dy_0. \quad (57)$$

Let's see a direct application of our results:

Example 4.1: Consider a network of $N = 2500$ mobile terminals distributed according to $f(x)$ in $[0, L]$ (for example, with $L = 5.6$ miles for WiMAX radius cell). We consider two base stations at position $BS_1 = 0$ and $BS_2 = L$ and $R = 1$. Then, the mobile association threshold (the boundary between both cells, i.e., the location at which the mobile terminals obtain the same throughput by connecting to any of both base stations) is reduced to find x such that the following equality holds:

$$(2^{N_1\theta} - 1)(1 + x^2)f(x) + 2^{N_1\theta}\theta \log 2 \left[x + \frac{x^3}{3} \right] = (2^{N_2\theta} - 1)(1 + (1 - x)^2)f(x) + 2^{N_2\theta}\theta \log 2 \left[\frac{4}{3} - 2x + x^2 - \frac{x^3}{3} \right] \quad (58)$$

Notice that this is a fixed point equation on x . If the mobile terminals are distributed uniformly, the optimal solution is given by $C_1 = [0, 1/2 L]$ and $C_2 = [1/2 L, L]$, which is the solution that Voronoi cells would give us and in that case the number of mobile terminals connected to each base station would be given by

$$N_1 = N_2 = 1250. \quad (59)$$

However, if the deployment distribution of the mobile terminals is more concentrated near BS_2 than BS_1 , consider for example $f(x) = 2x$, the optimal solution is given by $C_1 = [0, q]$ and $C_2 = [q, L]$ with $q = 0.6027 L$ and

$$N_1 = 908 \quad \text{and} \quad N_2 = 1592. \quad (60)$$

Notice that in the global optimization solution, the number of mobile terminals connected to BS_1 is smaller than the number of mobile terminals connected to BS_2 . However, the cell size is bigger.

V. RATE FAIR ALLOCATION POLICY

In this framework, we give the possibility to mobile terminals to associate to the base station they prefer in order to minimize their power cost function while maintaining, as quality of service measurement, an average throughput of θ .

As we presented in Section III, this problem is equivalent to

$$(RF) \quad \min_{C_i} \sum_{i=1}^K \iint_{C_i} \sigma^2(R^2 + d_i^2(x, y))^{\xi/2} f(x, y) dx dy. \quad (61)$$

Notice that this problem is equivalent to (P1) where the functions

$$F(d_i(x, y)) = (\sigma^2(R^2 + d_i^2(x, y))^{\xi/2}), \quad (62)$$

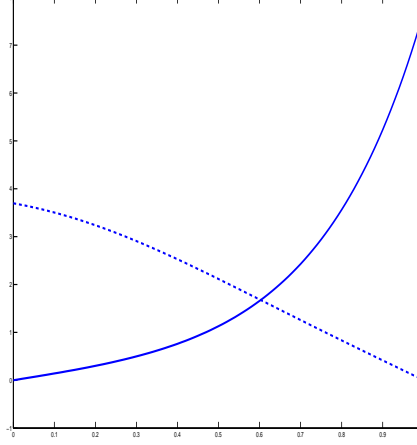
and $s_i \equiv 1$. The problem has then a solution given by

Proposition.- There exist a unique optimum given by

$$C_i = \left\{ x \in \mathcal{D} : \sigma^2(R^2 + d_i^2(x_0, y_0))^{\xi/2} \leq \sigma^2(R^2 + d_j^2(x_0, y_0))^{\xi/2}, \forall j \neq i \right\} \quad (63)$$

$$N_i = \iint_{C_i} f(x_0, y_0) dx_0 dy_0,$$

which is represented by the Voronoi cells.



(a) Equilibrium in Non-Uniform distribution

Fig. 9: Thresholds determining the cell boundaries (vertical axis) of the best equilibrium in terms of the SINR as a function of the location of BS_2 for BS_1 at position 0 when we consider a non-homogeneous distribution given by $f(x) = (L-x)/2L^2$. Example: equilibrium when the distribution of mobile terminals is given by $f(x) = 2x$ in the interval $[0, L]$ and the positions of the base stations are $BS_1 = 0$ and $BS_2 = L$.

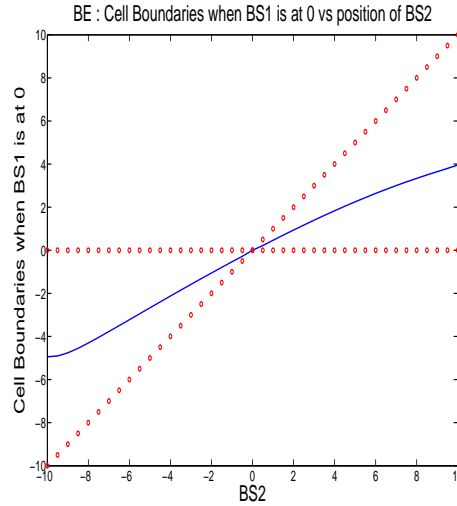


Fig. 10: Wardrop Equilibrium in the Non-homogeneous case: The network is deployed over the interval $[-10, 10]$ (one-dimensional case) presented here vertically. We find the threshold (solid line) determining the cell boundaries as a function of the base stations positions (dotted lines) by changing the position of one of them. BS_1 is fixed at position 0 and we change the position of BS_2 from -10 to $+10$. The deployment distribution of the mobile terminals is given by $f(x) = (L-x)/2L^2$.

A. Penalization function

Notice that the penalization function or the case when the number of mobile terminals is greater than the number of carriers available in the cell in the rate fair allocation policy case is equivalent to (P1) where the functions

$$F(d_i(x, y)) = (\sigma^2(R^2 + d_i(x, y)^2)^{\xi/2}). \quad (64)$$

$$s_i(N_i) = \kappa_i(N_i) = \begin{cases} 0 & \text{if } N_i \leq \text{MAX}_i, \\ \bar{\kappa}_i(N_i - \text{MAX}_i) & \text{if } N_i > \text{MAX}_i. \end{cases} \quad (65)$$

The problem has then a solution given by

Proposition.- There exist a unique optimum given by

$$\begin{cases} C_i = \{(x, y) : \sigma^2(R^2 + d_i^2(x_0, y_0))^{\xi/2} + s_i(N_i) + N_i \cdot s'_i(N_i) \\ \leq \sigma^2(R^2 + d_i^2(x_0, y_0))^{\xi/2} + s_j(N_j) + N_j \cdot s'_j(N_j)\} \\ N_i = \iint_{C_i} f(\omega, z) d\omega dz. \end{cases} \quad (66)$$

VI. PERFORMANCE GAP

As a second example consider again the case when the mobile terminals are uniformly distributed on $\mathcal{D} = [0, 1]$ but this time the two antennas are located at coordinates $x_1 = 0$ and $x_2 = 1$. Consider the case when $p = 1$ and

$$s_1(x) = 100 \quad \text{and} \quad s_2(x) = \begin{cases} 0 & \text{for } 0 \leq x \leq 0.999 \\ 1 & \text{for } 0.999 \leq x \leq 1. \end{cases}$$

Then the equilibrium cell configuration (C_1^E, C_2^E) is given by

$$C_1^E = \emptyset \quad \text{and} \quad C_2^E = [0, 1],$$

and the optimum cell configuration (C_1^*, C_2^*) is

$$C_1^* = [0, 0.001[\quad \text{and} \quad C_2^* =]0.001, 1].$$

The optimum is very unfair for mobile terminals living in the first cell C_1^* , who pay $x + 100$, whereas the other mobile terminals just pay the distance from 1. This is a toy example but it gives an idea of the performance gap between the centralized and the decentralized scenarios, also known as Price of Anarchy.

VII. NUMERICAL SIMULATIONS

In this section, we present several numerical results that validate our theoretical model.

A. One-dimensional case

We first consider the one-dimensional case and we consider a uniform distribution of mobile terminals in the interval $[-10, 10]$. We set the noise parameter $\sigma = 0.3$. In Fig. 7, we fix one base station BS_2 at position 0 and take as parameter the position of base station BS_1 . We consider as path loss exponent $\xi = 2$. Red lines shows the positions of the BSs. We are able to determine the cell boundary (solid blue curve) from BS_1 and BS_2 at different positions. In Fig. 11, we fix two base stations $\text{BS}_1 = -10$ and $\text{BS}_2 = 10$ and we take as parameter the position of base station BS_3 . Red lines shows the positions of the BSs. We determine the cell boundary (solid blue curve) from BS_1 and BS_3 and the cell boundary (dotted blue curve) from BS_2 and BS_3 .

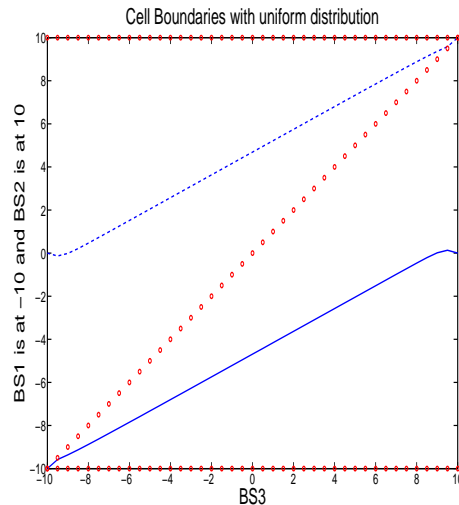


Fig. 11: Wardrop Equilibrium with Multiple Base Stations: The grid area network is the interval $[-10, 10]$ presented here vertically. We consider a uniform distribution of mobile terminals. We find the threshold (solid and dashed lines) determining the cell boundaries as a function of the base stations positions (dotted lines) by changing the position of one of them. BS_1 is fixed at position -10 and BS_2 is fixed at position 10 and we change the position of BS_3 from -10 to $+10$.

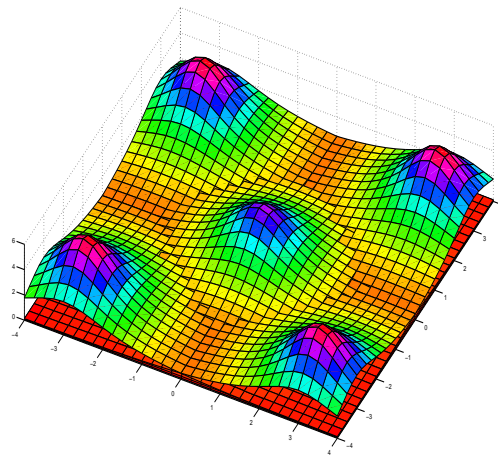


Fig. 12: Wardrop Equilibrium in the 2D case: The grid area network is the square $[-4, 4] \times [-4, 4]$. We set the noise parameter $\sigma = 0.3$ and we set four BSs at positions $BS_1 = (-3, -3)$ $BS_2 = (3, -3)$ $BS_3 = (3, 3)$ $BS_4 = (-3, 3)$ and one at the origin $BS_5 = (0, 0)$. We determine the cell boundaries (deep lines) for the uniform distribution of mobile terminals

B. Two-dimensional case: Uniform and Non-Uniform distribution of mobile terminals

We consider the two-dimensional case. We consider the square $[-4, 4] \times [-4, 4]$ and the noise parameter $\sigma = 0.3$. We set five base stations at positions $BS_1 = (-3, -3)$, $BS_2 = (3, -3)$, $BS_3 = (-3, 3)$, $BS_4 = (3, 3)$, and $BS_5 = (0, 0)$. We determine the cell boundaries for the uniform distribution of mobile terminals (see Fig. 13) and we compare it to the cell boundaries for the non-uniform distribution of mobile terminals given by $f(x, y) = (L^2 - (x^2 + y^2))/K$ where K is a normalization factor. The latter situation can be interpreted as the situation when mobile terminals are more concentrated in the center and less concentrated in suburban areas as in Paris, New York or London. We observe that the cell size of the base

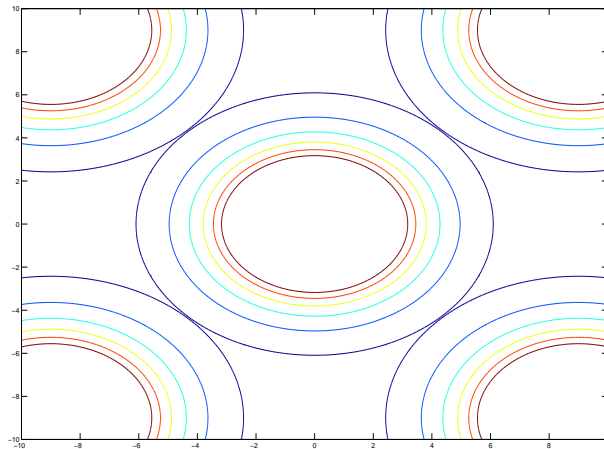


Fig. 13: 2D case: Cell contours of the best equilibrium with uniform distribution of mobile terminals.

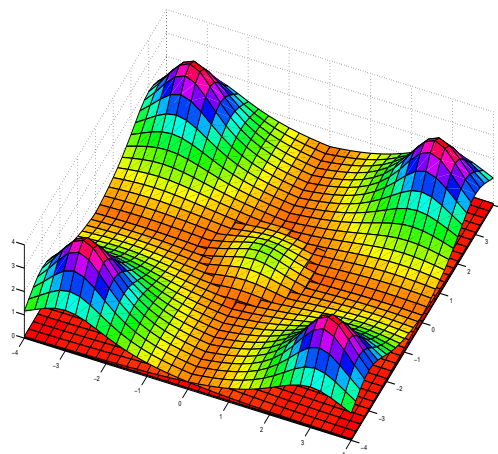


Fig. 14: Wardrop Equilibrium in 2D Non-Uniform Case: The grid area network is the square $[-4, 4] \times [-4, 4]$. We set the noise parameter $\sigma = 0.3$ and we set four BSs at positions $BS_1 = (-3, -3)$ $BS_2 = (3, -3)$ $BS_3 = (3, 3)$ $BS_4 = (-3, 3)$ and one at the origin $BS_5 = (0, 0)$. We determine the cell boundaries (deep lines) for the non-uniform distribution of mobile terminals given by $f(x, y) = (L^2 - (x^2 + y^2))/K$ where K is a normalization factor. The latter situation takes into account when mobile terminals are more concentrated in the center and less concentrated in suburban areas.

station BS_5 at the center is smaller than the others at the suburban areas. This can be explained by the fact that as the density of mobile terminals is more concentrated in the center the interference is greater in the center than in the suburban areas and then the SINR is smaller in the center. However the quantity of mobile terminals is greater than in the suburban areas.

VIII. CONCLUSIONS AND FUTURE PERSPECTIVES

In the present work, we have studied the mobile association problem in the downlink scenario. The objective is to determine the spatial locations at which mobile terminals would prefer to connect to a given base station rather than to other base stations in the network if they were offered that possibility (denoted decentralized scenario). We are also interested in the spatial locations which are more convenient from a

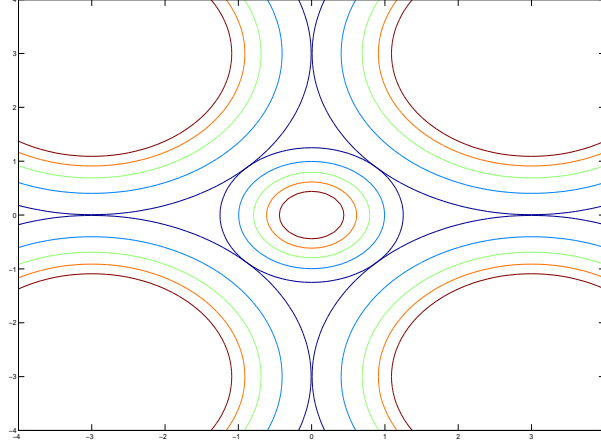


Fig. 15: 2D Non-Uniform: Cell contours of the best equilibrium with non-uniform distribution of mobile terminals.

centralized or from a network operator point of view. In both approaches, the optimality depends upon the context. In the considered cases, we consider the minimization of the total power of the network, which can be considered as an energy-efficient objective, while maintaining a certain level of throughput for each user connected to the network. We have proposed a new approach using optimal transport theory for this mobile association problems and we have been able to characterize these mobile associations under different policies.

The present work can be extended in several different directions. One of these possible directions is to study the price of anarchy between the centralized and decentralized scenario. As we presented in Section VI, the considered example give us an indication that the price of anarchy should be unbounded but currently we don't have precise bounds. The price of anarchy should be studied in both scenarios: the sum of a function and the multiplication. It should be interesting to study the application in the particular case when the network is an LTE network. Since our model is quite simplified in order to obtain exact solutions we could include the cases for the fading and shadowing effects. It is implicitly considered that the number of mobile terminals in the network is stationary, but since at different times of the day there are different number of mobile terminals, this management capabilities should be taken into account.

ACKNOWLEDGEMENTS

The authors would like to warmly thank Chloé Jimenez from *Université de Brest* for interesting discussions. The last author was partially supported by Alcatel-Lucent within the Alcatel-Lucent Chair in Flexible Radio at Supélec.

APPENDIX A

Consider the problem (P1)

$$\min_{C_i} \sum_{i=1}^K \iint_{C_i} \left[F(d_i(x, y)) + s_i \left(\iint_{C_i} f(\omega, z) d\omega dz \right) \right] f(x, y) dx dy, \quad (67)$$

where C_i is the cell partition of \mathcal{D} . Suppose that s_i are continuously differentiable, non-decreasing, and convex functions. The problem (P1) admits a solution that verifies

$$(S1) \begin{cases} C_i = \{ (x, y) : F(d_i(x, y)) + s_i(N_i) + N_i \cdot s'_i(N_i) \\ \leq F(d_j(x, y)) + s_j(N_j) + N_j \cdot s'_j(N_j) \} \\ N_i = \iint_{C_i} f(\omega, z) d\omega dz. \end{cases} \quad (68)$$

Proof.- The proof is based on Proposition 3.5 of Crippa et al. [6]. We include the proof for completeness and because part of it (mainly the existence of the solution) will be used in the proof of the following theorem. Notice that we have considered the case $F(d_i(x, y)) = |(x, y) - (x_i, y_i)|^p$, but this holds for any continuous function F .

From Section II, let us recall that Monge's problem can be stated as follows: given two probability measures, μ and ν , and a constant $p \geq 1$ we consider the minimization problem (denoted by $M_p(\mu, \nu)$):

$$M_p(\mu, \nu) := \inf \left\{ \left(\int_{\mathcal{D}} |x - T(x)|^p d\mu(x) \right)^{1/p} : T : \mathcal{D} \rightarrow \mathcal{D} \text{ Borel and such that } T\#\mu = \nu \right\}. \quad (69)$$

The relaxed formulation of Monge's problem (denoted by $W_p(\mu, \nu)$) can be stated as follows

$$W_p(\mu, \nu) := \inf_{\gamma \in \Pi(\mu, \nu)} \left\{ \left(\int_{\mathcal{D} \times \mathcal{D}} |x - y|^p d\gamma(x, y) \right)^{1/p} \right\}, \quad (70)$$

where $\Pi(\mu, \nu)$ is the set of probability measures such that $\pi_1(\gamma) = \mu$ and $\pi_2(\gamma) = \nu$ where π_1 is the projection on the first component and π_2 is the projection on the second component.

If the probability measure μ can be written as $d\mu = f(x) dx$ (i.e. it is absolutely continuous with respect to the Lebesgue measure), then the optimal values of both problems coincide $M_p(\mu, \nu) = W_p(\mu, \nu)$, and there exists an optimal transport map from μ to ν which is unique f -a.e. if $p > 1$. Another important characteristic of this relaxation is that it admits a dual formulation:

$$W_p^p(\mu, \nu) = \sup \left\{ \int_{\mathcal{D}} u d\mu + \int_{\mathcal{D}} v d\nu : \begin{array}{l} u(x) + v(y) \leq |x - y|^p \text{ for } \mu\text{-a.e. } x \text{ and } \nu\text{-a.e. } y \\ u \in L^1_{\mu}(\mathcal{D}), v \in L^1_{\nu}(\mathcal{D}) \end{array} \right\}. \quad (71)$$

Moreover, there exists an optimal pair (u, v) for this dual formulation, and when ν is an atomic probability measure (it can be written as $\nu = \sum_{i \in \mathbb{N}} b_i \delta_{y_i}$) the dual formulation becomes

$$W_p^p(\mu, \nu) = \sup \left\{ \int_{\mathcal{D}} u d\mu + \sum_{i \in \mathbb{N}} b_i v(y_i) : \begin{array}{l} u(x) + v(y_i) \leq |x - y_i|^p \text{ for } \mu\text{-a.e. } x \text{ and every } i \in \mathbb{N} \\ u \in L^1_{\mu}(\mathcal{D}), v \in L^1_{\nu}(\mathcal{D}) \end{array} \right\}. \quad (72)$$

There exists another interesting characteristic when one of the measures is absolutely continuous with respect to the Lebesgue measure and the other measure is an atomic measure. If the probability measure μ can be written as $d\mu = f(x) dx$ where f is a nonnegative function, $(y_i)_{i \in \mathbb{N}}$ is a sequence of points in the domain \mathcal{D} such that $\nu = \sum_{i \in \mathbb{N}} b_i \delta_{y_i}$, $(B_i)_{i \in \mathbb{N}}$ is a partition of the domain such that the map $T(x) = \sum_{i \in \mathbb{N}} y_i \mathbf{1}_{B_i}(x)$ is an optimal transport map from μ to ν , the pair (u, v) is a solution of the dual formulation (71), then

$$u(x) = \inf_{i \in \mathbb{N}} (|x - y_i|^p - v(y_i)) = \sum_{i \in \mathbb{N}} (|x - y_i|^p - v(y_i)) \mathbf{1}_{B_i}(x) \text{ for } f\text{-a.e. } x. \quad (73)$$

We also have a similar converse characteristic. If $(B_i)_{i \in \mathbb{N}}$ is a partition of \mathcal{D} , we set $b_i = \int_{B_i} f(x) dx$, $\nu = \sum_{i \in \mathbb{N}} b_i \delta_{y_i}$ and $T(x) = \sum_{i \in \mathbb{N}} y_i \mathbf{1}_{B_i}(x)$, and there exists two functions $u \in L^1_{\mu}(\mathcal{D})$, $v \in L^1_{\nu}(\mathcal{D})$ satisfying the condition (73), then T is optimal for $M_p(\mu, \nu)$ and the pair (u, v) is optimal for the dual formulation (71).

In order to prove the existence and uniqueness of the solution we need to consider the following: We denote by S the unit simplex in \mathbb{R}^k :

$$S = \left\{ c = (c_1, \dots, c_k) \in \mathbb{R}^k : c_i \geq 0, \sum_{i=1}^k c_i = 1 \right\}. \quad (74)$$

From Theorem 2.1, we deduce that

$$\inf_{\substack{(A_i)_{i \in \{1, \dots, k\}} \\ \text{partition of } \mathcal{D}}} \left\{ \sum_{i=1}^k \int_{A_i} \left[|x - x_i|^p + s_i \left(\int_{A_i} f(x) dx \right) \right] f(x) dx \right\} = \inf_{c \in S} \left\{ W_p^p(f, \sum_{i=1}^k c_i \delta_{x_i}) + \sum_{i=1}^k s_i(c_i) c_i \right\}. \quad (75)$$

The function $F : S \rightarrow \mathbb{R}$ defined by $F(c_1, \dots, c_k) = W_p^p(\mu, \sum_{i=1}^k c_i \delta_{x_i})$ is continuous and convex.

If the functions s_i are lower semi-continuous, then there exists an optimum and if in addition the maps $\eta_i(t) := th_i(t)$ are strictly convex, the optimum is unique.

We can then characterize the solution. If $(A_i)_{i=1, \dots, k}$ is an optimum, and s_i are differentiable in $]0, 1]$ and continuous in 0, then the following holds:

$$\begin{aligned} C_i = & \left\{ x \in \mathcal{D} : |x - x_i|^p + s_i(N_i) + N_i s'_i(N_i) \right. \\ & \left. \leq |x - x_j|^p + s_j(N_j) + N_j s'_j(N_j), \forall j \neq i \right\} \end{aligned} \quad (76)$$

$$N_i = \int_{C_i} f(x) dx. \quad (77)$$

APPENDIX B

Consider the problem (P2)

$$\text{Min}_{C_i} \sum_{i=1}^K \iint_{C_i} \left[F(d_i(x, y)) \cdot m_i \left(\iint_{C_i} f(\omega, z) d\omega dz \right) \right] \cdot f(x, y) dx dy, \quad (78)$$

where C_i is the cell partition of \mathcal{D} . Suppose that m_i are derivable. The problem (P2) admits a solution that verifies

$$(S2) \begin{cases} C_i = \{x : m_i(N_i) F(d_i(x, y)) f(x, y) + U_i(x, y) \\ \leq m_j(N_j) F(d_j(x, y)) f(x, y) + U_j(x, y)\} \\ U_i = m'_i(N_i) \iint_{C_i} F(d_i(x, y)) f(x, y) dx dy \\ N_i = \iint_{C_i} f(\omega, z) d\omega dz. \end{cases} \quad (79)$$

Proof.- The proof is similar to Appendix A. The problem (P2) can be rewritten as follows:

$$\inf \left\{ W_{c_i} \left(f dx, \sum_{i=1}^k c_i \delta_{x_i} \right) : c_i \geq 0, \sum_{i=1}^k c_i = 1 \right\}, \quad (80)$$

where

$$W_{c_i} \left(f dx, \sum_{i=1}^k c_i \delta_{x_i} \right) = \inf_{(A_i)_{i=1}^k} \left\{ \sum_{i=1}^k \int_{A_i} F(|x - x_i|) \cdot m_i(c_i) f(x) dx : \int_{A_i} f dx = c_i, (A_i)_{i=1}^k \text{partition de } \mathcal{D} \right\}. \quad (81)$$

Let γ be optimal for Monge-Kantorovich problem, then one can consider

$$\gamma(x, y) = \sum_{i=1}^k \gamma_i(x) \otimes \delta_{x_i}(y), \quad c_i = \gamma_i(\mathcal{D}), \quad (82)$$

where γ_i is a positive measure such that $\gamma_i \leq f dx$

The function $F : S \rightarrow \mathbb{R}$ defined by $F(c_1, \dots, c_k) = W_p^p(\mu, \sum_{i=1}^k c_i \delta_{x_i})$ is lower semi-continuous. Then there exists a solution and it is unique almost surely.

Let $(c_i)_{i=1}^k$ be a solution of (P2) and $(A_i)_{i=1}^k$ the associated optimal partition. Let us fix two indices i_0, j_0 and a point $x_0 \in A_{i_0}$. Let $\varepsilon \geq 0$. Let us consider the open ball of radius ε and center x_0 that we

denote as B_ϵ . We denote its measure as c_ϵ . We make a small variation of the optimal partition by taking from A_{i_0} the ball B_ϵ and adding it to A_{j_0} . Since the partition $(A_i)_{i=1}^k$ is optimal then

$$\begin{aligned} & \int_{A_{i_0}} F(|x - x_{i_0}|) h_{i_0}(c_{i_0}) f(x) dx + \int_{A_{j_0}} F(|x - x_{j_0}|) h_{j_0}(c_{j_0}) f(x) dx \\ & \leq \int_{A_{i_0} \setminus B_\epsilon} F(|x - x_{i_0}|) h_{i_0}(c_{i_0} - c_\epsilon) f(x) dx + \int_{A_{j_0} \cup B_\epsilon} F(|x - x_{j_0}|) h_{j_0}(c_{j_0} + c_\epsilon) f(x) dx, \end{aligned} \quad (83)$$

which is equivalent to

$$\begin{aligned} & \int_{A_{i_0}} F(|x - x_{i_0}|) (h_{i_0}(c_{i_0}) - h_{i_0}(c_{i_0} - c_\epsilon)) f(x) dx + \int_{B_\epsilon} F(|x - x_{i_0}|) h_{i_0}(c_{i_0} - c_\epsilon) f(x) dx \\ & \leq \int_{A_{j_0}} F(|x - x_{j_0}|) (h_{j_0}(c_{j_0} + c_\epsilon) - h_{j_0}(c_{j_0})) f(x) dx + \int_{B_\epsilon} F(|x - x_{j_0}|) h_{j_0}(c_{j_0} + c_\epsilon) f(x) dx. \end{aligned} \quad (84)$$

Dividing the previous equation by c_ϵ and taking the limit when $\epsilon \rightarrow 0$, we obtain

$$\begin{aligned} & \int_{A_{i_0}} F(|x - x_{i_0}|) h'_{i_0}(c_{i_0}) f(x) dx + F(|x - x_{i_0}|) h_{i_0}(c_{i_0}) f(x_0) \\ & \leq \int_{A_{j_0}} F(|x - x_{j_0}|) h'_{j_0}(c_{j_0}) f(x) dx + F(|x - x_{j_0}|) h_{j_0}(c_{j_0}) f(x_0). \end{aligned} \quad (85)$$

Reorganizing the terms we obtain the desired result.

REFERENCES

- [1] H. Hotelling, “Stability in competition,” *The Economic Journal*, vol. 39, no. 153, pp. 41–57, 1929. [Online]. Available: <http://www.jstor.org/stable/2224214>, Last visited: April 2010.
- [2] F. Plastria, “Static competitive facility location: An overview of optimisation approaches,” *European Journal of Operational Research*, vol. 129, no. 3, pp. 461 – 470, 2001. [Online]. Available: <http://www.sciencedirect.com/science/article/B6VCT-421TKPJ-1/2/47c0a2087d9d22e960343c3d931156b7>, Last visited: April 2010.
- [3] J. J. Gabszewicz and J.-F. Thisse, “Location,” in *Handbook of Game Theory with Economic Applications*, ser. Handbook of Game Theory with Economic Applications, R. Aumann and S. Hart, Eds. Elsevier, 1992, vol. 1, ch. 9, pp. 281–304. [Online]. Available: <http://ideas.repec.org/h/eee/gamchp/1-09.html>, Last visited: April 2010.
- [4] E. Altman, A. Kumar, C. K. Singh, and R. Sundaresan, “Spatial SINR games combining base station placement and mobile association,” in *INFOCOM*. IEEE, 2009, pp. 1629–1637. [Online]. Available: <http://dx.doi.org/10.1109/INFCOM.2009.5062081>, Last visited: April 2010.
- [5] G. Carlier, C. Jimenez, and F. Santambrogio, “Optimal transportation with traffic congestion and Wardrop equilibria,” *SIAM J. Control Optim.*, vol. 47, no. 3, pp. 1330–1350, 2008. [Online]. Available: <http://dx.doi.org/10.1137/060672832>
- [6] G. Crippa, C. Jimenez, and A. Pratelli, “Optimum and equilibrium in a transport problem with queue penalization effect,” *Adv. Calc. Var.*, vol. 2, no. 3, pp. 207–246, 2009. [Online]. Available: <http://dx.doi.org/10.1515/ACV.2009.009>, Last visited: April 2010.
- [7] C. Villani, *Optimal transport*, ser. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Berlin: Springer-Verlag, 2009, vol. 338, old and new. [Online]. Available: <http://dx.doi.org/10.1007/978-3-540-71050-9>, Last visited: April 2010.
- [8] —, *Topics in optimal transportation*, ser. Graduate Studies in Mathematics. Providence, RI: American Mathematical Society, 2003, vol. 58.
- [9] G. Monge, “Mémoire sur la théorie des déblais et des remblais,” *Histoire de l’Académie Royale des Sciences*, 1871.
- [10] L. V. Kantorovich, “On the transfer of masses,” *Dokl. Akad. Nauk.*, vol. 37, no. 2, pp. 227–229, 1942, translated in *Management Science*, Vol. 5, pp. 1-4, 1959.
- [11] G. Buttazzo and F. Santambrogio, “A model for the optimal planning of an urban area,” *SIAM J. Math. Anal.*, vol. 37, no. 2, pp. 514–530 (electronic), 2005. [Online]. Available: <http://dx.doi.org/10.1137/S0036141003438313>, Last visited: April 2010.
- [12] G. Carlier and I. Ekeland, “The structure of cities,” *J. Global Optim.*, vol. 29, no. 4, pp. 371–376, 2004. [Online]. Available: <http://dx.doi.org/10.1023/B:JOGO.0000047909.02031.ab>, Last visited: April 2010.
- [13] —, “Equilibrium structure of a bidimensional asymmetric city,” *Nonlinear Anal. Real World Appl.*, vol. 8, no. 3, pp. 725–748, 2007. [Online]. Available: <http://dx.doi.org/10.1016/j.nonrwa.2006.02.008>, Last visited: April 2010.
- [14] Y. Yu, B. Danila, J. A. Marsh, and K. E. Bassler, “Optimal transport on wireless networks,” Mar. 28 2007, comment: 5 pages, 4 figures. [Online]. Available: <http://arxiv.org/abs/physics/0703261>, Last visited: April 2010.
- [15] F. Baccelli and B. Blaszczyzyn, *Stochastic Geometry and Wireless Networks, Volume I - Theory*, ser. Foundations and Trends in Networking Vol. 3: No 3-4, pp 249-449, F. Baccelli and B. Blaszczyzyn, Eds. NoW Publishers, 2009, vol. 1, *Stochastic Geometry and Wireless Networks, Volume II - Applications*; see <http://hal.inria.fr/inria-00403040>. [Online]. Available: <http://dx.doi.org/10.1561/13000000006> <http://hal.inria.fr/inria-00403039/en/>, Last visited: April 2010.

- [16] Y. Brenier, “Décomposition polaire et réarrangement monotone des champs de vecteurs,” *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 305, no. 19, pp. 805–808, 1987.
- [17] H. Brezis, *Functional analysis, Sobolev spaces and partial differential equations*, ser. Universitext. New York: Springer, 2011.
- [18] G. S. Kasbekar, E. Altman, and S. Sarkar, “A hierarchical spatial game over licenced resources,” in *GameNets’09: Proceedings of the First ICST international conference on Game Theory for Networks*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 70–79.
- [19] J. Wardrop, “Some theoretical aspects of road traffic research,” *Proceedings of the Institution of Civil Engineers, Part II*, vol. 1, no. 36, pp. 352–362, 1952.